

Kalmár Péter

Adattudomány és „Big Data” technológia a controlling szolgálatában

A gyors fejlődés és a nagy mennyiségű, minőségi adat szinte minden gazdasági szereplő, így a vállalatok és vállalkozások, pénzügyi intézmények életére is jelentősen kihat. Az új technológia és az adattudomány katalizátora lehet olyan tudományágak tovább fejlesztésének, mint például a controlling. A controlling szolgálatába állított új megoldások, akár a tradicionálisan számviteli szemléletű és heurisztikus költségtervezési és elemzési eljárások számára is kínálhat alternatív, statisztikai alapú előrejelző és becslő módszereket. A szerző cikkében ezt az adattárolási és adatfeldolgozási módszert, annak lehetőségeit és a controlling rendszerhez való kapcsolódását ismerteti.

BEVEZETÉS

A 21. század második évtizedére az információ technológia (IT) fejlődésének üteme soha nem látott mértékkel gyorsult. Már nem csak az informatikai szakemberek de más ágazatok szakértői, valamint a mindennapi felhasználók számára is elérhetővé váltak az olcsó telekommunikációs és egyéb számítástechnikai hardverek, valamint a nyílt forrás-kódú szoftverek. A nyílt, megfizethető technológiák elterjedésének köszönhetően a technológia az emberi élet szinte minden területén jelen van ma már. A fejlődés újabb szintjét jelzi, hogy ezek a mindennapi eszközök okos, úgynevezett „smart” eszközökké válnak (ez a megnevezés első sorban az eszközök egymással való kommunikációjának képességére utal), amely az adatgyűjtés új dimenziót nyitja meg, egy részről a sebesség és adat mennyiség, más részről a környezetünk napi működésének új, eddig még nem rögzített területeinek tekintetében. A becslések szerint 2020-ra több mint 16 zettabyte¹ mennyiségű hasznos információt gyűjt össze az emberiség. [TURNER et al., 2014] Ennek az adatmennyiségnek a hatékony tárolása és feldolgozása az úgy nevezett „Big Data” technológia tárgya, míg az információ hatékony kinyerésével, hasznosításával és elemzésével a matematikai és statisztikai alapokra épülő adattudomány foglalkozik.

BIG DATA, ADATTUDOMÁNY ÉS CONTROLLING

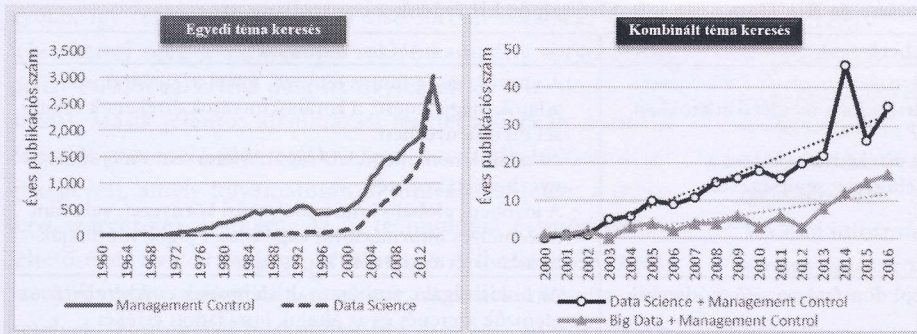
A SZAKIRODALOMBAN

Számos ágazatban a „Big Data” technológia egyaránt jelent lehetőséget és kihívásokat a szervezetek számára. A kifejezés valójában „a technológiai eszközökön keresztül (hitelkártyák, hűségpont-gyűjtő kártyák, közösségi média, internethez csatlakoztatott eszközök, vezeték nélküli szenzorok és jeladók, stb.) folyamatosan gyűjtött nagy mennyiségű adatra utal. Az adat jelentős része strukturálatlan, azaz semmilyen speciális, vagy előre meghatározott adat modellt nem követ.” [CHUA, 2013] A „Big Data” technológia köré épülő tudomány egy új szakág. Mi sem bizonyítja ezt

jobban, mint az a tény, hogy a technológia (nagy mennyiségű adat számítógépes klasztereken keresztül történő osztott tárolása és feldolgozása) első feljegyzett említése 2003-ra datált, amikor a Google programozói egy új fájlkezelési rendszerről szóló tudományos cikket publikáltak. [GHEMAWAT et al., 2003] Az új technológia, új eljárások kifejlesztését is igényelte, hogy a nagymennyiségű adatból információt, valamint gazdasági, tudományos vagy akár társadalmi értéket lehessen előállítani. A matematika, statisztika és programozás releváns részeinek kombinációja, az újonnan formálódó adattudomány („Data Science”) kínált metodológiai keretet ezen célok elérésére. Habár az adattudomány által alkalmazott matematikai és statisztikai eljárások és elméletek nagy része sokkal régebb óta létezik, mint maga az adattudomány, a gyakorlati alkalmazások, mint például a gépi tanulás („machine learning”) vagy a mesterséges intelligencia („artificial intelligence”) a „Big Data” technológiával párhuzamosan fejlődött.

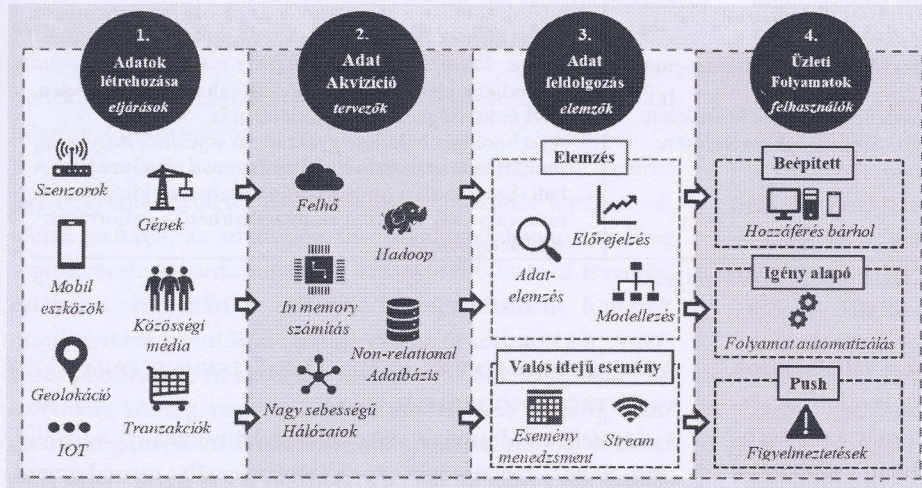
A „Big Data”-val, adattudománnyal és a menedzsment kontrollal kapcsolatos publikációk számának éves eloszlásával² kapcsolatban, a ProQuest³ átfogó publikációs adatbázis elemzése alapján elmondható, hogy az adattudomány szakirodalmi háttere, más szóval a publikációk éves száma (a széleskörű alkalmazási lehetőségeknek köszönhetően) már elérte a menedzsment kontrol publikációs szintjét (közel háromezer publikáció évente) egy évtizeden belül (lásd 1. számú ábra).

Ezzel egyidejűleg az adattudományt és a „Big Data” technológiát alkalmazó controlling szakterületet, azaz a három téma kombinált szakirodalmát, még új területnek számít és csak a korai fejlődés szintjét mutatja. Például a menedzsment kontrol és adattudomány kombinált szakirodalmának a 91%-a az elmúlt 10 évben került publikálásra (234 a 257 valaha regisztrált publikációból), míg a menedzsment kontrol és „Big Data” kombinált szakirodalmának a 93%-a (80 publikáció a 86 találatból). Az éves publikációs szám 2006 és 2016 között nem érte el a 50-et.



1. számú ábra: Az adattudomány és menedzsment kontrol kombinált publikációinak száma

Forrás: Saját szerkesztésű ábra a ProQuest adatbázis elemzése alapján



2. számú ábra: A „Big Data” Ökoszisztéma – az adatok keletkezésétől a döntésekig

Forrás: Saját szerkesztésű ábra (IoT a Dolgok Internetje (Internet of Things) kifejezésre utal) [STIRES, 2013] alapján

A „BIG DATA” ÖKOSZISZTÉMA

A „Big Data” terminológus nem csak a nagymennyiségű adatra, hanem az adatgyűjtés, tárolás és felhasználás különböző technológiai folyamataira, fázisaira, gyakorlatilag a Big Data Ökoszisztéma által leírt rendszerre is utal. A 2. számú ábrán látható a teljes rendszer az adatok keletkezésétől, azok döntésekben való felhasználásáig. Az ökoszisztéma négy fő fázisra épül:

1. Adatok létrehozása: Az elmúlt évtizedben, az internet alapú technológia gyors fejlődésének és elterjedésének köszönhetően, azon eszközök száma, amelyek az internethez vannak csatlakoztatva jelentősen megnövekedett. Egyes becslések szerint 2020-ra ez a szám, a dolgok internete (Internet of Things), elérheti a 30 milliárdot is, mivel nem csak új eszközök (telekommunikációs készülékek például), de hagyományos használati tárgyak (konyhai berendezések vagy éppen konnektorok is például) és gépek (személyautók, gyártógépek stb.) is adatforrássá válnak azáltal, hogy két irányú adatkommunikációra lesznek alkalmasak. Ezen képességükönél fogva az említett eszközök az emberi élet, de akár a vállalati folyamatok számos, eddig feltérképezetlen területeiről fognak részletes információt szolgáltatni. [MCEWEN AND CASSIMALLY, 2014]

2. Adat akvizíció: Az adatok akvizíciója a második eleme az ökoszisztémának, ez a szint az adatok összegyűjtését, tárolásáért, rendszerezéséért, infrastruktúrájáért és disztribúciójáért

felel. Három terület egyidejű fejlődése tette lehetővé az alrendszer létrejöttét: 1) A számítástechnikában használatos RAM valamint processzor teljesítmények növekedése, valamint ezen hardverek egységárak csökkenése; 2) a hatékony adatfeldolgozás menedzsment (klaszter alapú, disztribúciós feldolgozás, mint például Hadoop) eljárások térnyerése, valamint; 3) az adattárolási eljárások (például a felhő alapú módszerek) és a nagy sebességű hálózatok széleskörű elterjedése tette lehetővé, a nagy mennyiségű adatok valós-idejű rögzítését, tárolását, feldolgozását és disztribúcióját.

3. Adatfeldolgozás: A hagyományos adatelemzési eljárások a nagy mennyiségű adattal rendelkező, strukturálatlan adatbázisok esetén nem vagy csak alacsony hatásokkal alkalmazhatóak. Az új adattípusok, azok részletessége és időhorizontja a megnövekedett feldolgozó kapacitással kombinálva lehetővé teszik a valós-idejű adatfeldatárast, -elemzést és -modellezést, továbbá a pontos előjelzések készítését. Mindez kiegészítve a gépi tanulással és a mesterséges intelligenciával lehetővé tesz a teljesen automatizált esemény menedzsmentet (például a valós-idejű csalás vizsgálatot és prevenciót). Ez természetesen a matematika, statisztika és programozás kombinációját igényli, amelynek az adattudomány ad metodológiai keretet. [HEY et al., 2009].

4. Üzleti folyamatok és felhasználás: Az rendszer utolsó eleme az első három alrendszer kombinált eredményeire épít, az összegyűjtött és feldolgozott adatokból előállított hozzáadott értékre. Az adatok többféleképpen hasznosíthatók: egyrészt a vállalatok új bevételi forrásra tehetnek szert az adatok célzott tovább értékesítésével, másrészt pontosabb előjelzések és elemzések alapján jobb üzleti döntéseket hozhatnak a versenytársaknál azáltal, hogy az megfelelő információhoz a megfelelő időben (bárhol és bármikor) automatikusan hozzáférhetnek az üzleti döntéshozók. Egy másik, jelentős üzleti felhasználási terület az ügyfelek felé irányuló automatikus, releváns kommunikáció (például a Push Notification technológiával), mint például tranzakciók megerősítése vagy valós-idejű felmérések készítése.

Azok a gazdasági szervezetek, amelyek megtalálják az új információmennyiség hasznosításának megfelelő módját, képesek lesznek javítani a teljesítményüket, termelékenységüket, de akár új versenyelőnyre és bevételi forrásokra is szert tehetnek. Ugyanakkor a rosszul megtervezett és alkalmazott „Big Data” rendszerek, információ veszteséghez, biztonsági résekhez, valamint romló teljesítményhez vezethetnek. Az nem kérdés,

1. számú táblázat: Az adattudomány és Big Data adta lehetőségek és kihívások a controlling számára

Terület	Lehetőségek	Kihívások
Az adatállomány, mint üzleti érték	<ul style="list-style-type: none"> · Olyan eljárások kidolgozása, amelyek segítségével a vállalat képes lesz értékelni a meglévő adatokban rejlő lehetőségeket · Az adatok hozzáadott értékének növelése a minőségirányítás és felügyelet segítségével 	<ul style="list-style-type: none"> · A naprakészség fontos tényező, mivel az újabb típusú adatok megjelenése, a korábbi források értékének veszteségét eredményezheti. · Az adatok nem megfelelő hasznosítása szintén eredményezhet értékvesztést · A jövőbeni globális adatszabályozási környezet, valamint a személyes információval kapcsolatos jogok ronthatják az adatok értéktermelő képességét.
Adattudomány és Big Data a döntéshozatalban	<ul style="list-style-type: none"> · Specializált, valós-idejű döntéstámogató módszerek kidolgozása · Kollaboráció társosztályokkal annak érdekében, hogy új belső és külső adat hasznosítási módok kerüljenek azonosításra 	<ul style="list-style-type: none"> · Az önkiszolgáló, automatizált elemzések csökkenthetik az elemzők szerepét és az általuk hozzáadott értékét · A kulturális különbségek gátolhatják a (szervezetben belüli) globális adatmegosztást.
Adattudomány és Big Data a kockázatmenedzsmentben	<ul style="list-style-type: none"> · A kockázatelemzési kapacitás növelése, hogy egy pontosabb előrejelzési rendszer alakulhasson ki a tényezők komplex vizsgálata révén · Valós-idejű kockázat és csalás elemzés, „igazságügyi” számvitel kialakítása · Kockázat előrejelző modellek készítése hosszútávú pénzügyi instrumentumok és piacok értékelésére 	<ul style="list-style-type: none"> · Fontos, hogy a könnyen hozzáférhető és vizualizálható adatok korábban a korreláció és az ok-okozati kapcsolatok ne kerüljenek felcserélésre a kockázatelemzési eljárások során. · A prediktív elemzési technológiák változást jelentenek a ROI és költségkeret kalkulációkban is · A technológia fejlődése gyakran azt is jelenti, hogy a cég limitált tapasztalattal rendelkezik annak alkalmazásában. Ez a tanulási (próbálkozás-kudarcsiker) folyamat nem vagy csak nehezen számszerűsíthető a erőforrások allokációjakor.

Forrás: Saját szerkesztésű táblázat a 3.1-es ábra [CHUA, 2013, p.14.] alapján.

hogyan ezen technológiák szerepe jelentős lesz a vállalkozások életében, számos meglévő vállalkozás jelenleg is a felfedezés, rendszer és stratégia kialakítás fázisában tart már az új technológiával kapcsolatban, ugyanakkor a legtöbb új vállalkozásnak már magtevékenysége az információból történő gazdasági érték előállítása. Az adaptáció kérdése tehát a jövőben nem luxus kérdés lesz, hanem akár a cégek fennmaradásának alapfeltétele is lehet a modern piacokon. [CAVANILLAS et al., 2016]

Adattudomány: Amint azt eddig is láthattuk, a jelentősen megnövekedett információ mennyiség, választék és sebesség kombinációja új eljárásokat igényel annak érdekében, hogy a „Big Data” technológiából fakadó lehetőségeket az üzleti vállalkozások, kutatók és elemzők ki tudják aknázni. Az adattudomány („data-driven science”) egy interdiszciplináris tudomány terület, az olyan (matematikai és statisztikai) módszertannak, (programozási) folyamatoknak és rendszereknek a kapcsolatán alapuló tudományág, amelyek segítségével a különböző formában tárolt (strukturált és strukturálatlan) nagy mennyiségű adatokból tudományos, gazdasági vagy társadalmi érték állítható elő. [DHAR, 2013] Az adattudománynak számos alcsoportja ismert: feltáró adatelemzés, információ tervezés, interaktív adatvizualizáció, leíró statisztika, becslő eljárások, infografikák, stb. Az adattudósok az adatbányászat, tisztítás, feldolgozás és prezentáció területeinek szakemberei. Üzleti szempontból az adattudós azért, hogy képes legyen feltárni eddig rejtett üzleti megoldásokat, a készségek és szakmai ismeretek két csoportjával egyszerre kell rendelkezzen:

- technikai ismeretek (matematika, statisztika, SQL/python/C/Java programozás, adatbányászat, adatvizualizáció, Hadoop stb.), valamint
- üzleti készségek (elemzői és kommunikációs készségek, üzleti és ágazati ismeretek)

AZ ADATTUDOMÁNY ÉS A „BIG DATA” TECHNOLÓGIA A CONTROLLING SZOLGÁLATÁBAN

Az erőteljes információ központú adattudomány, valamint „Big Data” technológia és a jövőorientált menedzsment controlling koncepció egyidejű alkalmazása számos, jelenleg alkalmazott controlling eljárás fejlesztését teszi lehetővé. Az **1. számú táblázat** foglalja össze azokat a lehetőségeket és kihívásokat, amelyeket az új technológia jelent a pénzügyi, számviteli és controlling szakemberek számára. Az előrejelzések alapján 2020-ra a controlling rendszerek képesek lesznek valós-idejű adatelemzésre (például a terv-tény összehasonlítása, a tervtől való eltérés okainak valós-idejű feltárása és az automatikus újratervezés egyidőben), továbbá a tervezés és előrejelzés minőségének, frekvenciájának és pontosságának növelésére, valamint, hogy kritikus befektetés és stratégiai döntések meghozatalát támogassák a vállalati és üzleti információk széles körével.

Azok a controlling szakemberek, akik gyorsan alkalmazkodnak az új kihívásokhoz, valamint szakmai ismereteiket az adattudomány és „Big Data” technológia irányában kívánják kiterjeszteni, képesek lesznek arra, hogy erősítsék a szervezetek tervezési és elemzési képességeit, fejlesszék a döntéshozatali rendszereket, új piacokat azonosítsanak, valamint, hogy javítsák a működési hatékonyságot, amely végső soron magasabb profitszintet eredményez. [CHUA, 2013]

Alkalmazás a bank controlling területén: A pénzügyi szervezetek számára az információ már eddig is egy értékes eszköznek számított a működésben. A bankok jelenleg is a mindennapi működésük során számot fejtett statisztikai eljárást és modellt alkalmaznak például az Ügyfélkapcsolati Menedzsment (Customer Relationship Management – CRM) tevékenységük kapcsán, ahol a modellek segítségével azonosítják azokat az ügyfeleket, akik vásárlási hajlandósága na-

gyobb, így célzott marketinggel első sorban ezeket az ügyfeleket keresi meg a szervezet a reklámkampány során (ezzel pénzt és időt takarítva meg az ügyfélnek és a banknak egyaránt). Másik jelentős alkalmazási terület a kockázatelemzés és csalás prevenció (döntésfák valamint neurális hálózatok segítségével), amely folyamatosan új kihívásokat jelent a pénzügyi szervezetek számára. Az új technológia és módszertan lehetőséget nyit arra, hogy a pénzügyi és befektetési intézmények olyan szofisztikált, prediktív modelleket készítsenek, amelyek segítségével hosszútávú befektetések, piacok, pénzügyi instrumentumok (devizák, derivatívok és opciók) értékelése is lehetővé válik, amelyek eddig jelentős kockázatot jelentettek a kimenetel bizonytalansága miatt. Az új technológia arra ad lehetőséget eddig nem látott mennyiségű tényezőt, adatot, adattörténet vonjanak be az elemzők a prediktív modellekbe, ezáltal csökkentve azok egyszerűsítő hatását ezáltal növelve a pontosságot.

A fent említett három banki alkalmazási területen túl, a szerző egy negyedik területet is azonosított: a működés hatékonyságának javítását, az adattudományt alkalmazó bankcontrolling segítségével. A bankcontrolling témakörébe tartozó tervezés, értékelési, irányítási és információ menedzsment funkciók mindegyikére jelentős kihatással van az új technológia. A tervezés pontossága, frekvenciája javíthatóvá válik az új eljárások (prediktív költség vagy értékesítés modellezés) és az új adattömeg révén. Az értékelés során új módszertanok alkalmazásával lehetőség nyílik valós-idejű terv-tény elemzésre. Továbbá az új technológia az irányítás és információ menedzsment felgyorsulásához vezethet (bárhon, bármikor elérhető vállalati és részletes piaci, ügyfél stb. információ), valamint a döntési folyamatok leegyszerűsödhetnek. A gépi tanulás segítségével számos tervezési és értékelési folyamat válik automatikussá a jövőben, amely kihat a működési folyamatok hatékonyságára. A fejlődés jelenlegi ütemét követve a következő évtizedben kialakulhatnak az adattudományhoz és a „Big Data”-hoz kapcsolódó specifikus controlling területek (a project- és marketing controllinghoz hasonlóan).

LÁBJEGYZET

¹ zettabyte (ZB) = 1 billió gigabyte (GB) = 10^{21} byte (B)

² A keresés a következő publikációs formátumokra terjedt ki: Könyvek, Konferencia Kiadványok, Disszertációk, Műhelyviták, Szakmai folyóiratok. Nyelvi szűrés nem került alkalmazásra.

³ További információ a következő weboldalon érhető el: <http://www.proquest.com/>

FELHASZNÁLT IRODALOM

- CAVANILLAS, J.M.; CURRY, E.; WAHLSTER, W. (2016): New Horizons for a Data-Driven Economy. A roadmap for usage and exploitation of big data in Europe. Springer Open, USA, pp. 3-11., ISBN 978-3-319-21568-6
- CHUA, F. (2013): Big Data: its power and perils – Accountancy Futures Academy, The Association of Chartered Certified Accountants (ACCA), London, Last accessed on 28/02/2017 16:52 from: www.accaglobal.com/futures
- DHAR, V. (2013): Data science and prediction. Communications of the ACM, Vol (56), No (12), pp. 64-73. ISSN: 0001-0782
- GHEMAWAT, S.; GOBIOFF, H.; LEUNG, S.T. (2003): The Google File System, 19th ACM Symposium on Operating Systems Principles, Lake George, NY, October, 2003. Last accessed on 06/03/2016 from: <https://research.google.com/archive/gfs-sosp2003.pdf>
- HEY, T.; TANSLEY, S.; TOLLE, K. (2009): The fourth paradigm: Data-intensive Scientific Discovery, Microsoft Research, USA, pp. 15-22. ISBN: 978-0-9825442-0-4
- MCEWEN, A.; CASSIMALLY, H. (2014): Designing the Internet of Things, John Wiley and Sons Ltd. West Sussex, UK, pp. 7-20., ISBN: 978-1-118-43062-0
- STIRES, C. (2013): Big Data Ecosystem – From Data to decision, IDC's Asia/Pacific Big Data Conference 2013, IDC, Melbourne, Australia 03-Sep-2013, Last accessed on 05/03/2016 20:05 from: http://infographics.idc.asia/bigdata/Big_Data_ecosystem.asp
- TURNER, V.; GANTZ, J.E.; REINSEL, D.; MINTON, S. (2014): The digital universe of opportunities: rich data and the increasing value of the internet of things. IDC White paper. Last accessed on 05/03/2016 from <http://idcdocserv.com/1678>

CONTROLLER INFO

Tudományos szakfolyóirat – jog, vezetői számvitel, pénzügy, gazdaság

V. évfolyam, 2. szám

2017. II. negyedév

Adattudomány és „Big Data” technológia a controlling szolgáltatásban

Belső minőségirányítási rendszer bevezetésének
pénzügyi és controlling szempontú vizsgálata egy
működő vállalkozás példáján

Észak-alföldi élelmiszer-kiskereskedelmi vállalkozások
pénzügyi helyzetének elemzése diszkriminancia-analízis
segítségével

